

PROTEIN AND GLYCOPROTEIN CHARACTERISATION BY MASS SPECTROMETRY

AUTHOR: HOWARD R. MORRIS, EMERITUS PROFESSOR OF BIOLOGICAL CHEMISTRY AND SENIOR RESEARCH INVESTIGATOR, DEPARTMENT OF MOLECULAR BIOSCIENCES, IMPERIAL COLLEGE, LONDON, UK; FORMER CHAIRMAN OF THE M-SCAN GROUP OF COMPANIES; SGS CONSULTANT.

Proteins are polymers made up of amino acids linked via their alpha amino and carboxyl groups respectively, creating planar interconnecting amide bonds. The order of the twenty naturally occurring amino acids in the polymer is called the amino acid sequence or primary structure, and all other levels of protein structure, including secondary (alpha helix, pleated sheet or random coil), tertiary (three dimensional) structure, or quaternary structure (interaction with a ligand or subunit in space) are fundamentally determined by the unique amino acid sequence of the protein. Determining the sequence of an unknown protein has therefore been the holy grail of protein chemistry ever since Fred Sanger's pioneering work on the structure elucidation of Insulin – the first protein sequence determination - in the 1950s¹.

For the following two decades classical Edman sequencing was often the method of choice for protein analysis, although in the background, considerable research effort was under way examining the potential of new sequencing technologies based on mass spectrometry (MS). This work began in the late 1950's with the analysis of simple synthetic di- and tri- peptides of known sequence to determine whether the fragmentation patterns observed in the mass spectrometer could be used to deduce sequence. Despite the demonstration of significant potential in this method by the Scandinavian, Russian, American and French groups, including a spectacular success on a microbial depsipeptide peptidolipid by an Anglo-French team, all attempts to sequence protein-derived peptides by mass spectrometry had failed.

The solution to the problem came from the realisation in 1968 that in purifying the peptides produced from the necessary step of digesting these giant polymer molecules, vast quantities of organic impurities were being added to the small amounts of natural protein-de-

rived peptides, masking their appearance at the mass spectrometer detector.

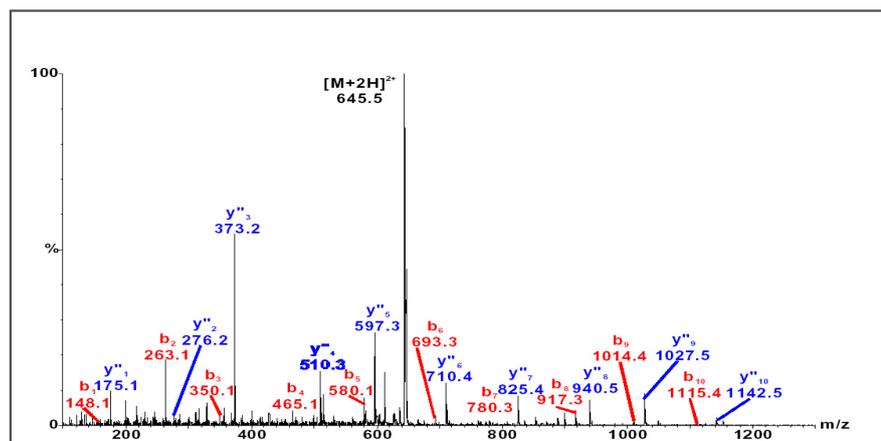
The impurities were leaching unnoticed from the ion exchange columns – the most advanced high resolution purification method of the day – as substituted styrene multimers. Since synthetic peptides were purified by crystallisation, these problems had not been encountered on the molecular standards used earlier in the proof-of-principle studies. A novel final purification step involving fractional distillation using a source temperature gradient was all that was then required to produce the first successful sequence analysis of protein-derived peptides in 1968^{2,3}, in a study of the unknown sequence of silk fibroin (later found to be over 1 million M.W.) from the silk worm *Bombyx mori*.

The mechanistic principles of how peptides fragment in a mass spectrometer were fully determined over the following decade, and the sequence determination process is illustrated using a modern proteomics example in Figure 1A and 1B, showing MS/MS fragmentation of an unknown sequence from a fucosyl transferase on a Q-TOF tandem qua-

drupole orthogonal acceleration time of flight instrument^{4,5}, and the interpretation of the data ((b)) to build a sequence for database searching of genomic and protein libraries. Basically, fragment ions created by cleavage across the peptide bond can hold amino-terminal (b ion) or carboxy-terminal (y ion) sequence information and the sequence is built up by counting the mass differences between the fragment ions – corresponding to the masses of the amino acid residues, and extrapolating between them.

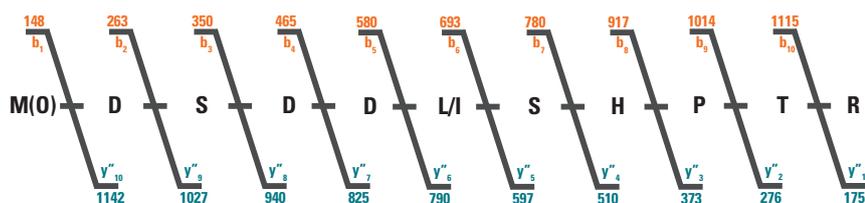
A serendipitous event in the early silk fibroin work led to another key advance in sequencing technology, the novel idea that mass spectrometry could allow the simultaneous sequencing of peptide mixtures ("mixture analysis"). Prior to that, major efforts were expended on purifying single pure peptides from digest mixtures (actually an important pre-requisite in classical sequencing strategies), until a chance MS analysis of a sample which was thought to be pure, but in fact was a mixture of very similar sequences, led to the discovery that simultaneous sequencing of multicomponent mixtures was possible by mass spectrometry, thus

FIGURE 1A: PROTEOMICS APPLICATIONS



DE NOVO SEQUENCING : STUDY OF DAUGHTERS OF $[M+2H]^{2+}$ 645.5 (FTase) NANOSPRAY- ES CAD MS/MS ON Q-TOF

FIGURE 1B: PROTEOMICS APPLICATIONS



INTERPRETATION OF CAD MS/MS DERIVED DATA FROM $[M+2H]^{2+}$ 645.3 NANOSPRAY- ES CAD MS/MS ON Q-TOF

removing the rate-limiting step (the purification process) in sequencing strategies. Following that breakthrough, numerous *de novo* sequencing studies on important biological problems established and then reinforced the credibility of the method, to the point that some twenty years later it almost completely replaced classical Edman as the method of choice for protein sequencing. Those founding studies included the structure elucidations of bioactive molecules including the brain's own opiate Enkephalin⁶ the first of the Endorphins, of Adipokinetic hormone⁷, of Aplysia cardio active peptide⁸, of human

calcitonin gene-related peptide CGRP⁹, of Antithrombin variants leading to deep vein thrombosis (DVT)¹⁰, together with the protein structures of Chloramphenicol Acetyltransferase¹¹ and of the vitamin K-dependant Prothrombin structure crucial to the initial steps in blood coagulation¹².

During the early development of MS protein strategies involving the mixture analysis protocol, it was recognised that using specific enzymes to cut the peptide backbone led to a set of resulting peptides whose molecular masses were

characteristic of the particular protein under study. This observation, in turn, led to the concept of mass mapping (mass fingerprinting)¹³ for protein identification and for the monitoring of any changes in anticipated structure, for example arising from post-translational events such as glycosylation, or mutational events in the gene coding for the protein. In situations where a protein sequence is known, mass mapping data alone, rather than actual sequence data, may be sufficient evidence of the correctness or otherwise of a predicted structure. At this point, in the early 1980's, the newly formed company M-SCAN was able to assist the fledgling biotechnology industry in the characterisation of their genetically - engineered recombinant protein biopharmaceutical products, including Interferons and protein hormones¹⁴.

M-SCAN was the first company to pioneer the offering of these commercial services to assist the biotechnology and pharmaceutical industries and a simple example of that early work is shown in Figure 2A and 2B for the fusion product of Urogastone. The manufacturer had been unable to confirm the correct amino acid sequence of the product by Edman degradation which had failed after the first ten to fifteen residues, whereas the peptide map produced by M-SCAN, by MS analysis of the tryptic digest, clearly showed the correct masses for all the predicted peptides, including the C-terminal one at m/z 633, providing evidence for the correctness of the structure. Peptide mapping is now recognised as an important step in the characterisation of any biopharmaceutical product and is included in the ICH Q6B guidelines issued by the global regulators through the International Committee on Harmonisation.

Many bioactive proteins found in nature are in fact glycoproteins i.e. proteins in which various amino acids are covalently linked to sugar residues. The first mass

FIGURE 2A: PEPTIDE MAPPING/FINGERPRINTING

AN EARLY BIOTECH APPLICATION

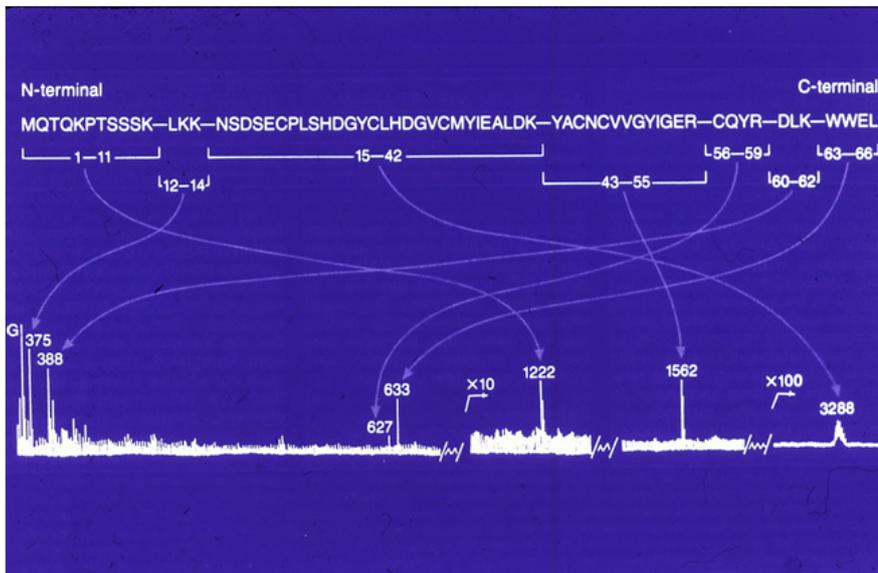
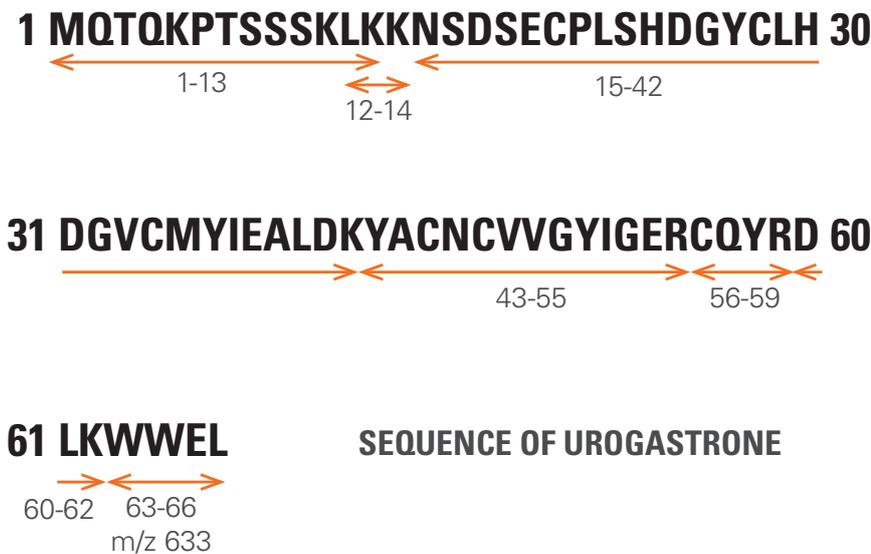


FIGURE 2B: MASS SPECTROMETRY IN PROTEIN ANALYSIS



spectrometric study of such a molecule resulted from the work of a University of California, Davis scientist, Professor Bob Feeney, who was fascinated by why Antarctic fish swimming under the sea ice don't freeze to death, since they contain circulating blood similar to humans! Prof. Feeney isolated blood fractions possessing antifreeze properties, which was not just salt (which can of course depress the freezing point of water), but rather a complex set of glycoproteins. In a collaboration using a variation of the newly developed mass spectrometric protein strategies, the study of Feeney's molecules resulted in the finding of some very unusual polypeptide structures, decorated at every third residue (Threonine) with a di-saccharide galactosyl (1-3) N-Acetyl galactosamine O-linked to the amino acids^{15,16}. The glycosyl hydroxyls in this type of structure are now thought to play a role in hydrogen bonding to nano-ice crystals, thus preventing their continuing growth and endowing antifreeze properties on the glycoproteins. Some of the MS data used for the interpretations, together with the structure discovered for fraction AF8 are shown in Figure 3A, 3B and 3C. Following this work, further mass spectrometric studies on other unknown molecules have validated the power of MS for the analysis of protein glycosylation^{17,18,19}, and today MS is the method of choice for glycoprotein characterisation and features in the ICH Q6B guidelines for this type of biopharmaceutical analysis. Mass spectrometry is used to study both O- and N-linked glycosylation, for the mapping of the glycoforms, as illustrated in Figure 4, at each substitution site on the protein and to characterize and identify which amino acid residues carry the glycosyl substituents.

FIGURE 3A: AF8:THE PEPTIDE STRUCTURE
DIRECT PROBE MS OBTAINED BY THE FRACTIONAL DISTILLATION TECHNIQUE

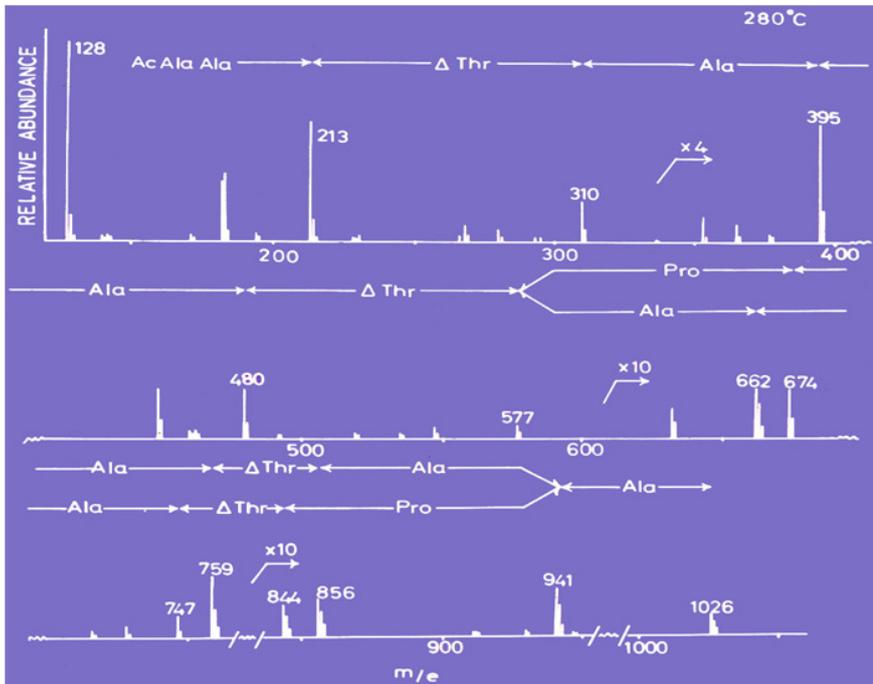
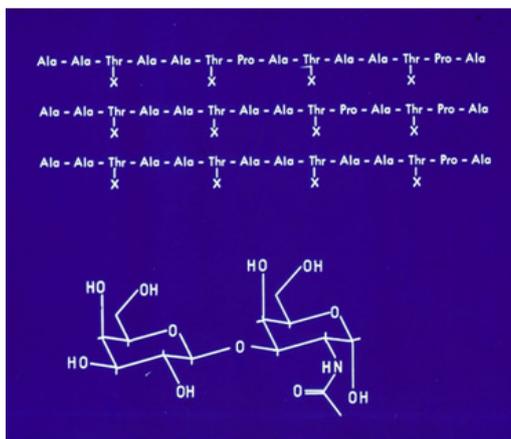
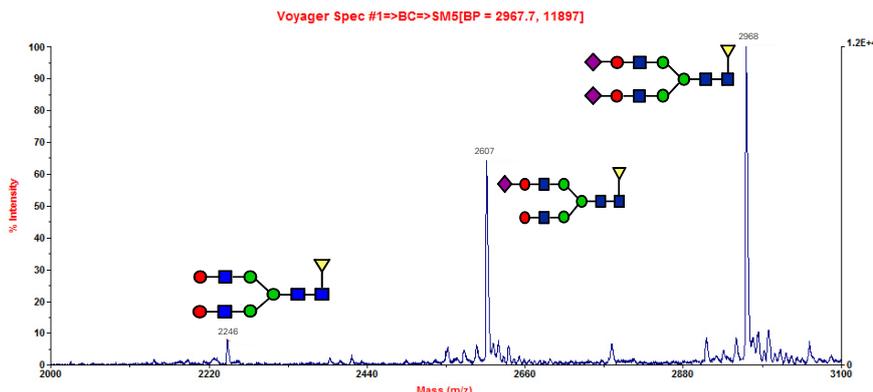


FIGURE 3C: THERE IS ANTIFREEZE IN THE BLOOD!

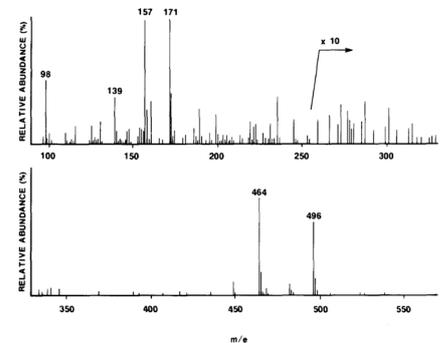


**ANTARCTIC FISH
 AF8 GLYCOPROTEIN**

FIGURE 4: MALDI MS CHARACTERISATION OF N-LINKED GLYCANS :



**FIGURE 3B: THE
 CARBOHYDRATE STRUCTURE**

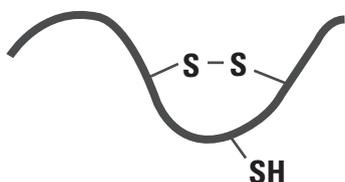


**AMMONIA CHEMICAL IONISATION (CI)
 MASS SPECTRUM OF
 GLYCOPROTEIN AF8**

The formation of the correct disulphide bridge configuration, linking the Cysteine thiol groups in a the protein, is another important post-/co-translational event in the production of an active, properly folded protein or glycoprotein product. Incorrectly folded proteins will almost certainly lack their biological efficacy and may be antigenic. An analysis of the disulphide bridge configuration is therefore a significant step in an overall characterization package. This analysis, however, is not without difficulty, even using the advanced mass spectrometric strategies which we specifically developed for this task²⁰. Practical problems include the possibility of disulphide bridge scrambling due to oxido-reductive events, and also the sheer complexity of possible structures in cases where the protein Cysteine content is high causes interpretation problems of its own. Figure 5A, 5B and 5C illustrates the strategy developed for disulphide bridge analysis, where it should be appreciated that even for a simple protein containing three Cysteine's (as shown) the experimental strategy would need to differentiate between four possible structures (three different disulphide configurations and one all-reduced structure). Figures 5B and 5C show how subsequent experimental steps including reduction and/or MS/MS can be used to support an initial interpretation.

FIGURE 5A: CHARACTERISATION OF S-S BRIDGES

DISULPHIDE BRIDGED PROTEIN

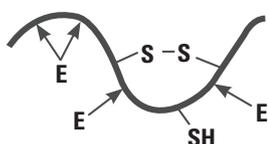


A strategy for disulphide bridge analysis based on mass mapping/fingerprinting and the recognition that different disulphide bridged peptides will have different masses

Morris H.R. and Pucci, P. *Biochem. Biophys. Res. Commun.* **126**, 1122-1128 (1985)

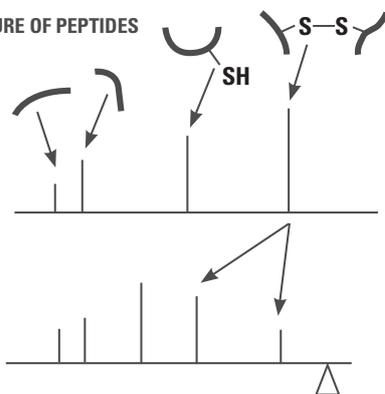
FIGURE 5B: CHARACTERISATION OF S-S BRIDGES

DISULPHIDE BRIDGED PROTEIN



ENZYMATIC/
CHEMICAL DIGESTION

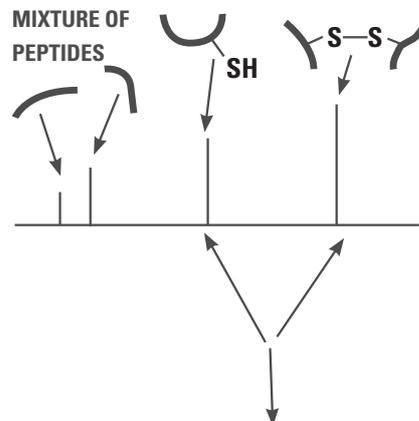
MIXTURE OF PEPTIDES



IDENTIFICATION BY MS FOLLOWED BY REDUCTION AND FURTHER MS

FIGURE 5C: CHARACTERISATION OF S-S BRIDGES

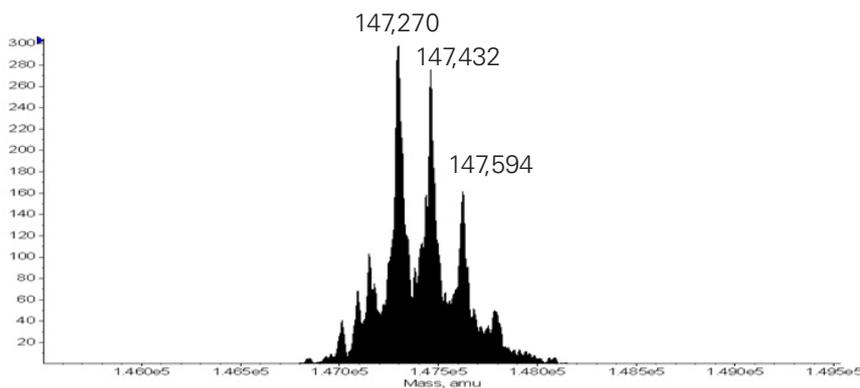
MIXTURE OF PEPTIDES



ADDITIONAL VERIFICATION BY MS/MS OR EDMAN DEGRADATION

This short paper gives the essential background to an understanding of biopharmaceutical characterisation by mass spectrometry and its importance in accelerating the processes leading to the approval by the regulator of well-characterised new drug products, including monoclonal antibodies, hormones and cytokines, and the new generation of Biosimilars and Biobetters resulting from patent expiration of the originator products. Subsequent newsletters from SGS M-Scan experts will provide more detail of the glycosylation, disulphide bridge assignment and monoclonal antibody specifics of the MS characterisation protocols offered by SGS M-SCAN to serve its clients' needs.

FIGURE 6: CHARACTERISATION OF BIOPHARMACEUTICALS
ON-LINE LC/ES-MS OF AN INTACT MONOCLONAL ANTIBODY (mAb)



The unrivalled experience of M-SCAN in protein/glycoprotein analysis is of great benefit to the clients, since every type of product is different with respect to its physico-chemical properties, including for example solubility or hydrophobicity (leading to adsorption losses), and with respect to the complexity of structural characterisation required. The new generation of complex monoclonal antibody drugs (mAbs) illustrate quite well the

range of problems facing the mass spectroscopist, with molecular weights in the 150,000 Dalton range, with numerous inter-and intra-disulphide bridges, with glycosylation, usually involving range of glycoforms as seen in Figure 6, and other structural peculiarities including blocked N-termini or C-terminal truncations. The challenge of a full characterisation is a formidable one.

REFERENCES

1. Sanger, F. and Tuppy, H. *Biochem. J.* **49**, 463-481 (1951)
2. Morris, H.R., Geddes, A.J. & Graham, G.N. *Biochem. J.*, **111**, 38 (1969)
3. Geddes, A.J. et al *Biochem. J.* **114**, 695-702 (1969).
4. Morris, H.R. et al *Rapid Communications in Mass Spectrometry* **10**, 889-896 (1996).
5. van der Wel, H. et al *Journal of Biological Chemistry* **276**, 33952-33963 (2001).
6. Hughes, J. et al *Nature* **258**, 577-579 (1975).
7. Stone, J.V. et al *Nature*, **263**, 207-211 (1976).
8. Morris, H.R. et al *Nature* **300**, 643-645 (1982).
9. Morris, H.R. et al *Nature* **308**, 746-748 (1984).
10. Erdjument, H. et al *Journal of Biological Chemistry* **262**, 13381-13384 (1987).
11. Shaw, W.V. et al *Nature* **282**, 870-872 (1979).
12. Morris, H.R. et al *Biochem. J.* **153**, 663-679 (1976).
13. Morris, H.R. Panico, M. & Taylor, G.W. *Biochemical and Biophysical Research Communications* **117**, 299-305 (1983).
14. Greer, F.M. & Morris, H.R. *Protein Engineering* **1**, 243 (1987).
15. Morris, H.R. et al *J. Biol. Chem.* **253**, 5155-5161 (1978).
16. Morris, H.R. and Thompson, M.R. in *Advances in Mass Spectrometry in Biochemistry and Medicine* (Frigerio, A. and Castagnoli, N. eds) **1**, 239-245 (1976)
17. Robb, R. et al *PNAS* **81**, 6486-6490 (1984).
18. Morris, H.R. et al *Journal of Biological Chemistry* **271**, 32159-32167 (1996).
19. Tissot, B. et al *FEBS Letters* **583**, 1728-1735 (2009).
20. Morris, H.R. & Pucci, P. *Biochemical and Biophysical Research Communications* **126**, 1122-1128 (1985).

To receive future articles on current trends and regulatory updates, subscribe to SGS' Life Science News at www.sgs.com/lss_subscribe

CONTACT INFORMATION

EUROPE

BELGIUM

+32 10 42 11 11
be.pharmaqc@sgs.com

FRANCE (PARIS)

+33 1 41 06 95 93
fr.pharmaqc@sgs.com

FRANCE (POITIERS)

+33 (0) 5 49 57 04 04
clinicalresearch@sgs.com

GERMANY (BERLIN)

+49 30 3460 7500
de.pharmaqc@sgs.com

GERMANY (FREIBURG)

+49 761 6116 7760
de.pharmaqc@sgs.com

GERMANY (TAUNUSSTEIN)

+49 6128 744 245
de.pharmaqc@sgs.com

SWITZERLAND (GENEVA)

+41 22 794 8374
pharmaqc@sgs.com

UK (WOKINGHAM)

+44 (0) 1189 896940
pharmaqc@sgs.com

ASIA

INDIA

+91 44 2254 2601
in.pharmaqc@sgs.com

SINGAPORE

+65 677 53 034
sg.pharmaqc@sgs.com

CHINA

+86 21 6115 2197
cn.pharmaqc@sgs.com

TAIWAN

+886 2 2299 3279 ext 2500
tw.pharmaqc@sgs.com

NORTH AMERICA

CANADA

+ 1 905 364 3757
ca.pharmaqc@sgs.com

USA (FAIRFIELD, NJ)

+ 1 888 747 8782
us.pharmaqc@sgs.com

USA (LINCOLNSHIRE, IL)

+1 847 821 8900
us.pharmaqc@sgs.com

USA (WEST CHESTER, PA)

+ 1 610 696 8210
us.pharmaqc@sgs.com

WWW.SGS.COM/PHARMAQC